Validation of the Treatment Motivation Scales for forensic outpatient treatment (TMS-F)

Drieschner, Klaus H.; Boomsma, Anne

Published in:
Assessment

DOI:
10.1177/1073191107311651

Publication date:
2008

Validation of the Treatment Motivation Scales for Forensic Outpatient Treatment (TMS-F)

Klaus H. Drieschner and Anne Boomsma

Assessment 2008; 15; 242 originally published online Feb 1, 2008;
DOI: 10.1177/1073191107311651

Validation of the Treatment Motivation Scales for Forensic Outpatient Treatment (TMS-F)

Klaus H. Drieschner
Forensic Psychiatric Center Oldenkotte, Netherlands

Anne Boomsma
University of Groningen, Netherlands

*The validity of the Treatment Motivation Scales for outpatient offender treatment (TMS-F), a self-report questionnaire with scales for the motivation of patients to engage in the treatment and six cognitive and emotional determinants of this motivation, is evaluated in two studies. In Study 1 (*N *= 620), the* construct validity *of the TMS-F is investigated applying a multitrait-multimethod design with a therapist-rating instrument as the criterion method. All scales were found to have adequate convergent validity and acceptable discriminant validity. In Study 2 (*N *= 328), the* criterion validity *of the TMS-F is addressed. Applying covariance structure analysis, the instrument is found to predict therapist ratings of the treatment engagement of patients to a substantial degree. Treatment engagement is best predicted by the scales for Motivation to Engage in the Treatment and Perception of the Suitability of the Treatment and not by the scales for Distress and the Perceived Legal Pressure.*

Keywords: offender treatment; treatment motivation; treatment engagement; test construction; validation; multitrait-multimethod

The importance of the patient's treatment motivation is increasingly emphasized in the field of offender treatment (e.g., Barbaree, 2005; Howells & Day, 2007; McMurran, 2002; Tierney & McCabe, 2002). This has several reasons. First, it has become widely recognized that the patient's responsivity to the treatment is critical for the treatment outcome, and that the treatment motivation is an important aspect of this responsivity (Andrews, Bonta, & Wormith, 2006; McGuire, 1995; Serin & Kennedy, 1997). Second, treatment motivation and engagement are considered important for determining the risk of future violence (e.g., Webster, Douglas, Eaves, & Hart, 1997). Finally, measuring treatment motivation is important in outcome research. More than once, the failure to include this variable as a covariate has limited the interpretability of otherwise well-designed outcome studies (Falshaw, Friendship, Travers, & Nugent, 2003; Marques, Wiederanders, Day, Nelson, & Van Ommeren, 2005).

Despite the widely recognized importance of the patient's motivation in the field of offender treatment, there is a dearth of measurement instruments for this variable. Available instruments are either conceptually questionable, such as the University of Rhode Island Change Assessment (McConnaughy, Prochaska, & Velicer, 1983), applicable

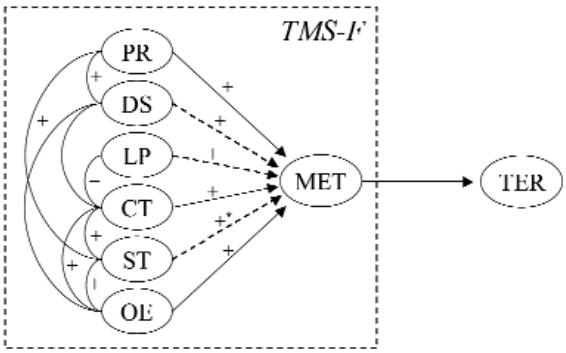
Klaus H. Drieschner, Forensic Psychiatric Center Oldenkotte, Netherlands; Anne Boomsma, Department of Sociology, University of Groningen, Netherlands. Klaus H. Drieschner is now at Trajectum Hoeve Boschoord, Boschoord, Netherlands. The article is based on the first author's doctoral dissertation. Correspondence concerning this article should be addressed to Klaus H. Drieschner, Reiderlandlaan 1, 9727 DR Groningen, Netherlands; e-mail: K.Drieschner@hoeveboschoord.dji.minjus.nl.

Assessment, Volume 15, No. 2, June 2008 242-255
DOI: 10.1177/1073191107311651

to only a particular category of offenders, such as the Anger Readiness to Change Questionnaire (Williamson, Day, Howells, Bubner, & Jauncey, 2003) and the Texas Christian University Motivation Scales (Knight, Holcom, & Simpson, 1994), or impractical due to a very long administration time, such as the offender adaptation of the Personal Concerns Inventory (Sellen, McMurran, Cox, Theodosi, & Klinger, 2006). The Dutch Treatment Motivation Scales for Forensic outpatient treatment (TMS-F), an 85-item self-report instrument which contains seven content scales and a 15-item social desirability scale, was developed as an alternative which does not have these shortcomings (Drieschner & Boomsma, 2008).

The TMS-F was developed according to the construct method (Jackson, 1971). This deductive approach to questionnaire design starts with the specification of a nomological network, which consists of the definitions of the concepts to be measured and a theory for the relationships between these concepts. According to the theory underlying the TMS-F, which was first described by Drieschner, Lammers, and Van der Staak (2004), the patient's Motivation to Engage in the Treatment depends on six interrelated internal determinants, labeled Problem Recognition, Distress, Perceived Legal Pressure, Perceived Costs of the Treatment, Perceived Suitability of the Treatment, and Outcome Expectancy. On its part, the Motivation to Engage in the Treatment determines the patient's treatment engagement (see Figure 1). Descriptions of the content domains of the TMS-F concepts, on the basis of which the TMS-F items were generated, are provided in Table 1. The TMS-F and its theoretical basis were put to a test in two forensic outpatient samples employing confirmative factor analysis and structural equation modeling (Drieschner & Boomsma, 2008). In both samples, the hypothesized factorial structure of the TMS-F was supported, and five of the seven content scales turned out to be essentially one-dimensional. The Motivation to Engage in the Treatment and Perceived Costs of the Treatment scales represented three and two dimensions, respectively, which were sufficiently correlated to use the sum scores of the whole scales. Cronbach's alpha values for the TMS-F content scales ranged between .81 and .91. In addition, most hypothesized relationships among the seven scales were supported by the data. About 60% of the variance of Motivation to Engage in the Treatment was accounted for by the internal determinants. However, in contrast with expectations, Distress and Perceived Legal Pressure had no effect on the Motivation to Engage in the Treatment, and the effect of Problem Recognition was only modest.

FIGURE 1
Originally Hypothesized and Empirically Supported Models for the Prediction of Treatment Engagement by the TMS-F Scales



Note: PR = Problem Recognition; DS = Distress; LP = Perceived Legal Pressure; CT = Perceived Costs of the Treatment; ST = Perceived Suitability of the Treatment; OE = Outcome Expectancy; MET = Motivation to Engage in the Treatment; TER = Treatment Engagement Rating scale. + = Positive Association; – = Negative Association; ------- = originally hypothesized but not empirically supported.
*redundant in multivariate prediction of MET despite substantial correlation.

Because the TMS-F had to be broadly applicable among Dutch forensic outpatient populations, which include individuals at risk of offending, who receive voluntary preventive treatment, the items do not refer to specific problem areas, such as substance abuse, anger problems, sex offending, or offending in general. The TMS-F was designed for the assessment of the patient's motivation during the course of the treatment. Because several items refer to the patient's perception of the treatment, the instrument is not suitable to assess initial treatment motivation.

In the present article, the validity of the TMS-F is evaluated on the basis of correlations with external criterion measures. Which kind of validity is required depends on the intended applications of an instrument and the type of conclusions one wants to be able to draw from the test scores. According to Cronbach (1990), "validation is inquiry into the soundness of the interpretations proposed for the scores from a test" (p. 145). The six scales for internal determinants of Motivation to Engage in the Treatment—Problem Recognition, Distress, Perceived Legal Pressure, Perceived Costs of the Treatment, Perceived Suitability of the Treatment, and Outcome Expectancy—must enable an analysis of motivational problems of individual patients. For this purpose, the exact meaning of the test scores must be clear and thus *construct validity* is of primary importance. Because the relevance of the motivation to engage in the

TABLE 1
Description of the Concepts Underlying the TMS-F Scales

Concept	*Acronym*	*Components*
Motivation to engage in the treatment	MET	1. Commitment for treatment engagement and session attendance[a] 2. Commitment for treatment completion[a] 3. Readiness to make sacrifices (money, emotional burden, lifestyle)[a] 4. Readiness for disclosure 5. Readiness to make treatment efforts between sessions
Problem recognition	PR	1. Recognition that one must change to prevent recidivism 2. Acknowledgment that one needs treatment to achieve this change
Distress	DS	1. Distress resulting from symptoms, demoralization, sense of inferiority, social problems, worry about the future, and dissatisfaction with life
Perceived legal pressure	LP	1. Belief that lack of engagement will lead to expulsion from treatment 2. Belief that drop-out or expulsion will lead to a legal sanction 3. Perception of the possible legal sanction as threatening
Perceived costs of the treatment	CT	1. Aversiveness (e.g., psychological burden of the treatment, narcissistic injury of being in need of psychological help) 2. Sacrifices (time, money, lifestyle, reputation among peers)
Perceived suitability of the treatment	ST	1. Perceived suitability of the method and rationale of the treatment 2. Agreement about the goals of the treatment 3. Satisfaction with the therapist
Outcome expectancy	OE	1. Expectancy that one will be able to finish the treatment 2. Expectancy that the treatment leads to the intended behavior change 3. Expectancy that the treatment will result in a better life

a. Items for these three components constitute one Motivation to Engage in the Treatment subscale.

treatment is mainly based on its relationship with the behavior of treatment engagement, the *predictive validity* is pivotal for the Motivation to Engage in the Treatment scale. In the sequel, two validation studies are reported, in which the construct validity of all TMS-F scales (Study 1) and the predictive validity of the Motivation to Engage in the Treatment scale (Study 2) are put to the test. Research concerning the social desirability scale will be reported in another publication but the most important findings are summarized in the General Discussion section (see also Drieschner, 2005).

STUDY 1

Introduction

An essential part of the construct validation of a measure is the investigation of its convergent and discriminant validity. The most rigorous method for this purpose is the multitrait-multimethod (MTMM) methodology, in which the convergent and discriminant validity can be addressed simultaneously. Convergent validity is inferred from correlations with other measures for the same trait, the monotrait-heteromethod correlations. Discriminant validity requires that the monotrait-heteromethod correlations of a scale exceed the heterotrait-heteromethod and heterotrait-monomethod correlations. Furthermore, construct validity is supported if the patterns of trait-method correlations of different methods converge.

Validation of a new instrument with a MTMM design faces two major obstacles. First, validated measures for the traits of interest, which could be used as a criterion method, are typically not available. After all, this is the reason why the new instrument is developed. In the present study, a therapist-rating instrument for the same concepts underlying the TMS-F scales was used as criterion method. The second obstacle is the lack of consensus about how to analyze MTMM data (Kenny & Kashy, 1992), and the fact that available methods often do not yield interpretable solutions (e.g., Marsh & Grayson, 1995). In the present study, the Correlated Trait-Correlated Method Minus One (CT-C[M-1]) model (Eid, 2000; Eid, Lischetzke, Nussbeck, & Trierweiler, 2003) is applied, which provides a remedy for some technical and interpretational problems of earlier methods.

Method

Participants. The present sample consisted of 620 patients from a larger sample (N = 754; Drieschner & Boomsma, 2008), for whom therapist ratings on the criterion measure were available. The 620 participants in this

study received treatment in 1 of 10 forensic outpatient treatment centers in the Netherlands. In these centers court mandated treatment is provided to offenders with psychiatric or personality problems, as well as voluntary treatment to individuals who are considered at risk of (re-)offending by mental health care providers, general practitioners, or the police (Plemper, 2001; Salize & Dressing, 2005). The primary purpose of all treatments is prevention of offending. The predominant treatment approach in all treatment centers is cognitive-behavioral. In most cases (78%), the treatment consisted of between one and four therapy sessions a week, often in a combination of individual and group therapy. About 15% of the patients followed treatment programs with an intensity of two or more whole days a week. In 7% of the cases, patients had less than one treatment session a week. At all treatment sites, psychotherapists, psychologists, social workers, and psychiatrists were involved in the treatments. Some sites employed additional professionals such as art, music, psychomotor, and marital therapists. Inclusion criteria, which correspond with the boundaries of the target population of the TMS-F, were (a) at least three previous therapy sessions, (b) at least one treatment session scheduled within three weeks (to exclude individuals with incidental appointments for the purpose of support of control), (c) termination of the treatment not planned within 8 weeks, and (d) the ability to read and understand simple Dutch sentences. Reliable information concerning refusal rates was not available for the whole sample, but evidence from one treatment center ($N = 200$) with a 90% participation rate indicated that only about 5% of those patients who were invited refused to participate. Characteristics of the present sample are summarized in Table 2.

Procedure. The study was approved by a recognized ethical committee according to APA ethical standards. The data collection took place in two waves from June 2001 to January 2002 ($N = 285$) and from September 2002 to September 2003 ($N = 335$). Patients who met the inclusion criteria were invited by their therapists. When informed consent was obtained, the therapists handed out the TMS-F form and entered the patient's name on a registration list. Patients on that list were not approached again. The patient completed the TMS-F during or after the treatment session and handed it over in a closed envelope to the therapist, the first author, or a research assistant. Therapists completed the therapist-rating instrument, and provided the information summarized in Table 2 on a separate form. Neither the therapist nor the patient was informed about the scores provided by the respective other person. At 7 of the 10 sites ($N = 488$), patients were offered €5 for participation.

Measures. For a description of the TMS-F see the Introduction section. Only 8 Problem Recognition and 13 Motivation to Engage in the Treatment items were used because for part of the sample no data were available for four items, which were added to the instrument in the final version (see Drieschner & Boomsma, 2008). In the subsample of 335 cases for which data on all items were available, the correlations between the 8-item and 9-item Problem Recognition scales and between the 13-item and 16-item Motivation to Engage in the Treatment scales were both .99.

The second measure for the TMS-F concepts was a 20-item therapist-rating instrument consisting of a 5-item Motivation to Engage in the Treatment scale and 2-item or 3-item scales for the six internal determinants of Motivation to Engage in the Treatment. The item stems contained—in slightly reworded form—the descriptions of the 19 components of the TMS-F concepts (see Table 1), on the basis of which the TMS-F items were generated. For example, the stems of the three Outcome Expectancy items were (a) The patient is confident to complete the treatment, (b) The patient expects to change because of the treatment, and (c) The patient expects that the treatment will result in a better life. Although the Distress scale of the TMS-F contained only a single component (see Table 1), two Distress items were included in the therapist-rating instrument, (a) The patient suffers from their own problem behavior or psychological problems, and (b) The patient suffers from the consequences of the problem behavior, including worry about the future. The raters received the written instruction: *Please indicate to what extent the descriptions given in the items below apply to the patient by checking one of the five response options*. Ratings were provided on 5-point scales with endpoints labeled *Not* and *Strongly*. The first two Perceived Legal Pressure items had an additional *Does not apply* response option for voluntary patients, which was coded as the lowest level of Perceived Legal Pressure. Because the scores correlated almost perfectly, the mean score of these two items was used. The raters received no further instruction or training. No prior psychometric evaluation of the instrument had taken place. Construct validation addresses the question whether an instrument measures the intended concept. The fact that the item stems of the therapist-rating instrument contain the very descriptions of the intended content domains of the TMS-F scales makes this instrument a natural criterion for the construct validation of the TMS-F.

Information about the patients' age, gender, ethnicity, type of offense, legal status, time in treatment, and psychopathology was obtained on a separate form. For each of the variables, therapists had to choose between the predefined alternatives reported in Table 2.

TABLE 2
Sample Characteristics in Study 1 and Study 2

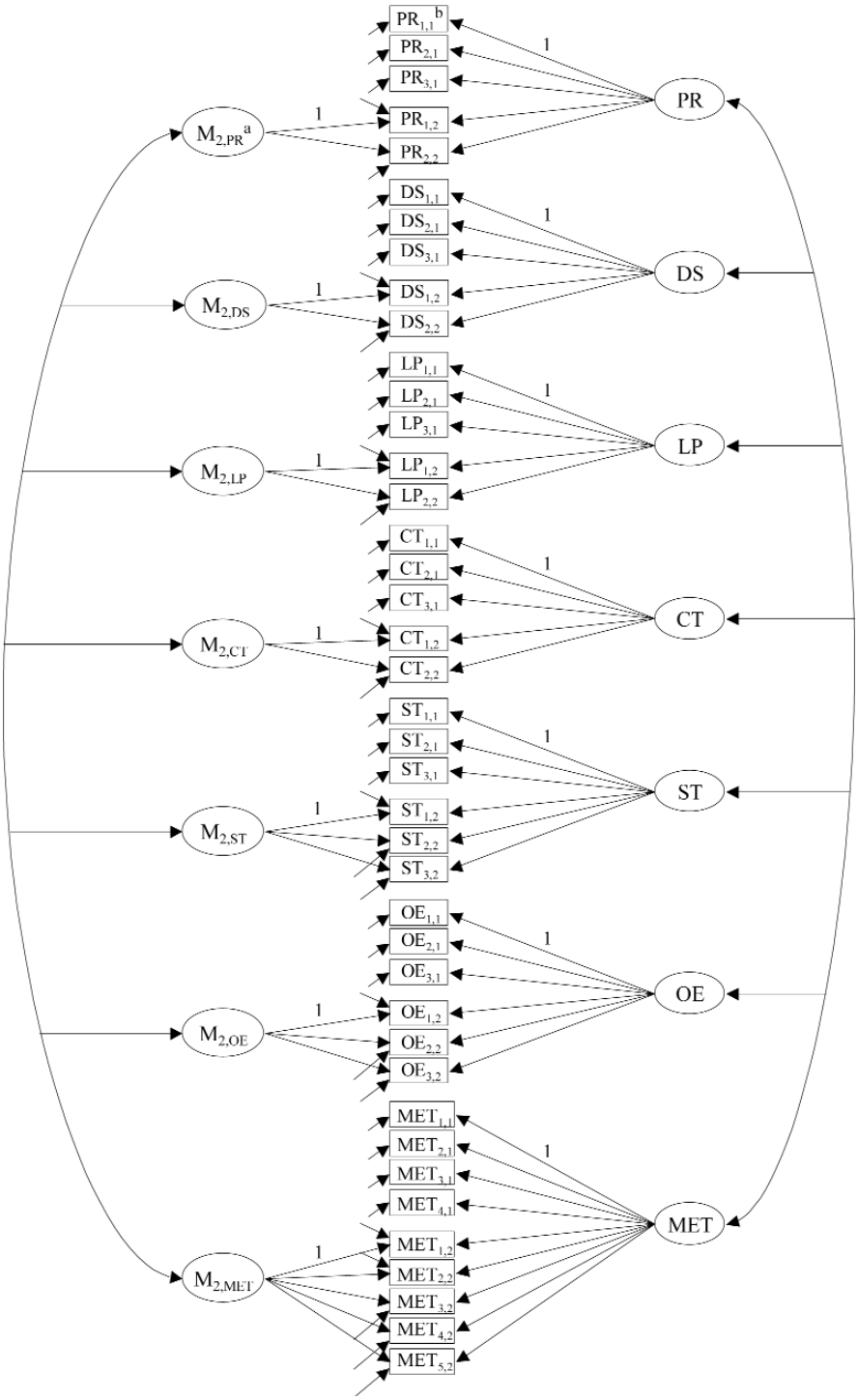
Study 1	*Study 2*	*(N = 620)*	*(N = 328)*
Age	Average years (*SD*)	37.5 (12.0)	36.1 (11.1)
Gender	Male	90.3	89.9
Ethnicity	Both parents Dutch nationality	91.3	89.6
	At least one parent Asian (including Turkey), African, or South American nationality	7.6	9.1
	Other	1.1	1.3
Type of offense	Sexual offense		
	Victim < 16 years	31.9	22.6
	Victim ≥ 16 years	7.7	8.2
	Violent offense (no sexual or material motive)	32.3	37.8
	Miscellaneous (e.g., burglary, stalking, arson)	12.9	14.9
	No known offense	15.2	16.5
Legal status	Court mandated	54.1	52.8
	Voluntary	42.2	44.2
	Not yet sentenced	3.7	3.0
Time in treatment	3 months or less	39.7	37.5
	4-9 months	29.0	29.0
	10-18 months	19.0	20.4
	More than 18 months	12.3	13.1
Psychopathology	Axis I		
	No axis I disorder	39.3[a]	38.3[b]
	Psychotic disorder	5.9[a]	8.3[b]
	Addiction	16.2[a]	21.9[b]
	Other axis I disorder	44.1[a]	38.8[b]
	Axis II		
	Substantial characteristics of:		
	Cluster B personality disorder	46.1[a]	46.8[b]
	Cluster A or C personality disorder	33.5[a]	36.3[b]
	No personality disorder	22.5[a]	19.2[b]

NOTE: With the exception of "age" all entries indicate percentages.
a. $N = 614$.
b. $N = 322$.

Data Analysis. To reduce the complexity of the MTMM model and to obtain approximately equal numbers of indicators for both methods, the TMS-F items were aggregated into item-parcels, and parcel-scores instead of item-scores were used in the MTMM analysis. Four parcels for the 16-item Motivation to Engage in the Treatment scale, and three parcels for each of the 9-item internal determinant scales were formed, employing a procedure described by Little, Cunningham, Shahar, and Widaman (2002, p. 166). With this procedure, items are assigned to parcels in such a way that each parcel consists of a one-dimensional set of items, and that the parcels for the same scale have similar loadings on their common factor.

A CT-C(M-1) model with multiple indicators for each trait-method unit (Eid et al., 2003) was fitted to the data (see Figure 2). While in a conventional Correlated Trait-Correlated Method (CT-CM) model the variance of each indicator variable is decomposed into a trait, a method, and an error component, in a CT-C(M-1) model no method factor is linked to the indicators of the method of primary interest, in the present study the TMS-F. By fixing the loading of one TMS-F item-parcel for each scale to the value of one, the trait factors were defined as the true-score variables of these item-parcels. Consequently, the method factors are residual factors, which represent the deviations of the true scores of the therapist ratings from the values that would be expected on the basis of the trait factors. Because the method factors are defined as residual factors, they are uncorrelated with the trait factors of the same trait. This allows for the decomposition of the variance of each indicator variable into a trait-specific, a method-specific, and an error component. The trait-specific variance components represent the variance explained by the trait factors and are reflected by the *consistency coefficients*. Because of the asymmetry of the CT-C(M-1) model, the consistency coefficients of both methods have different interpretations. Because the trait factors are defined by one of the TMS-F parcels for the respective traits, the consistency coefficients for the TMS-F simply reflect the common variance of the TMS-F parcels in the respective scale. More important are the consistency coefficients of the therapist ratings, which reflect the proportion of the

FIGURE 2
Multiple-Indicator Correlated Trait-Correlated Method Minus One [CT-C(M-1)] Model with the TMS-F as Comparison Standard



NOTE: Correlations between Trait- and Method factors belonging to different traits are not depicted to avoid overloading the figure. PR = Problem Recognition; DS = Distress; LP = Perceived Legal Pressure; CT = Perceived Costs of the Treatment; ST = Perceived Suitability of the Treatment; OE = Outcome Expectancy; MET = Motivation to Engage in the Treatment.
a. Method factor of Method 2 (therapist rating) for the trait indicated in the subscript.
b. First subscript: indicator of trait-method combination, second subscript: method (1: TMS-F, 2: therapist rating).

TABLE 3
Observed Heterotrait-Monomethod Correlations Among Scales for the TMS-F and Therapist Ratings

	PR	*DS*	*LP*	*CT*	*ST*	*OE*	*MET*
PR	1	0.58	–0.27	0.26	0.56	0.57	0.48
DS	0.54	1	–0.23	0.12	0.34	0.30	0.36
LP	–0.09	–0.16	1	–0.29	–0.15	–0.20	–0.17
CT	0.09	–0.25	–0.13	1	0.35	0.34	0.36
ST	0.27	–0.07	0.08	0.46	1	0.69	0.60
OE	–0.06	–0.42	0.14	0.44	0.63	1	0.63
MET	0.14	–0.19	0.02	0.42	0.48	0.59	1

NOTE: PR = Problem Recognition; DS = Distress; LP = perceived Legal Pressure; CT = perceived Costs of the Treatment; ST = perceived Suitability of the Treatment; OE = Outcome Expectancy, MET = Motivation to Engage in the Treatment. Below diagonal: correlations among composite scores of TMS-F parcels. Above diagonal: correlations among composite scores of therapist-ratings.

TABLE 4
Observed Heteromethod Correlations Between TMS-F Scales and Therapist Ratings

	Therapist Ratings						
TMS-F	*PR*	*DS*	*LP*	*CT*	*ST*	*OE*	*MET*
PR	.39	.27	–.24	.11	.24	.21	.16
DS	.33	.40	–.25	.06	.10	.02	.01
LP	–.24	–.17	.61	–.24	–.11	–.14	–.15
CT	.07	–.03	–.14	.32	.23	.23	.21
ST	.16	.07	–.11	.19	.40	.39	.30
OE	–.03	–.10	.05	.08	.23	.30	.23
MET	.14	.04	.04	.15	.29	.32	.31

NOTE: PR = Problem Recognition; DS = Distress; LP = perceived Legal Pressure; CT = perceived Costs of the Treatment; ST = perceived Suitability of the Treatment; OE = Outcome Expectancy; MET = Motivation to Engage in the Treatment.

variance of the therapist ratings accounted for by the corresponding TMS-F scales. These coefficients constitute the basis from which the convergent validity coefficients of the TMS-F scales are derived. The *method-specificity coefficients* reflect the proportions of the reliable variances of the therapist ratings, which are not explained by the trait factors. The sum of the consistency and the method-specificity coefficients equals the *reliability coefficient.* Finally, the error components are those parts of the variance of indicator variables, which are accounted for by neither the trait nor the method factors. Estimates of the trait and the method components of the true-score variables are obtained by dividing the consistency and method-specificity coefficients of the observed variables by the corresponding reliability coefficients. Because only the reliable variance is decomposed, the consistency and method-specificity coefficients of the true-score variables add up to the value of one. The purpose of using several indicators for each trait-method combination is to separate measurement error from trait-specific method effects. Once this is accomplished, aggregation formulas (see Eid et al., 2003) are applied to obtain reliability, consistency, and method-specificity coefficients at the level of the scales. Of primary interest are the square roots of the true-score consistency coefficients of the therapist-rating scales, which represent the latent correlations between the two methods for the same traits and are typically reported as convergent validity coefficients.

Maximum likelihood estimation with mean- and variance-based adjustment for model fit (MLMV) as implemented in the Mplus software (Muthén & Muthén, 2004) was applied to analyze the CT-C(M-1) model. Model fit was evaluated by the adjusted chi-square statistic, χ^2_{MLMV}, the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the Tucker-Lewis Index (TLI), and the Comparative Fit Index (CFI). Values lower than .08 for the RMSEA and SRMR, and higher than .90 for the TLI and CFI are conventionally interpreted as indicating an acceptable fit. More recent criteria (Hu & Bentler, 1999) suggest that the risk of wrongly accepting the present model is negligible when SRMR < .055 or when SRMR < .10 in combination with either RMSEA < .06, TLI > .95, or CFI > .96.

Before the more sophisticated CT-C(M-1) model was analyzed, the correlations among the observed scores of the various trait-method combinations were evaluated, as recommended by Marsh and Grayson (1995).

Results

About 0.8% missing values were imputed by a simple but effective two-way imputation method that takes into account the respondent's average score on the other items of the scale and the average score of all respondents on the item for which the value is missing (Sijtsma & Van der Ark, 2003).

Observed Correlations Among Trait-Method Units. As can be seen in Table 3, the general patterns of heterotrait-monomethod correlations for the TMS-F (below diagonal) and the therapist ratings (above diagonal) showed marked similarities but also some differences. For both methods, the strongest associations were found between Problem Recognition and Distress and between Perceived Suitability of the Treatment, Outcome Expectancy, and Motivation to Engage in the Treatment. Furthermore, Perceived Costs of the Treatment was substantially correlated with Perceived Suitability of the Treatment, Outcome Expectancy, and Motivation to Engage in the Treatment. Differences were the association of Problem Recognition with Outcome Expectancy, which was found only for the therapist ratings, and the associations of Distress with Perceived Costs of the

TABLE 5
Variance Components of TMS-F Scales and Therapist Ratings Computed from a Correlated Trait-Correlated Method Minus One (CT-C[M-1]) Model

			Observed Scores			*True Scores*		
	λ_{Trait} *Range*	λ_{Method} *Range*	*Reliability*	*Consistency*	*Method-Specificity*	*Consistency*	*Method-Specificity*	*Latent Correlation*
PR-TMS	.80-.82		.85	.85		1		
PR-ThR	.35-.44	.72-.80	.85	.18	.67	.21	.79	.46
DS-TMS	.87-.91		.92	.92		1		
DS-ThR	.31-.42	.56-.77	.75	.17	.58	.23	.77	.48
LP-TMS	.83-.87		.91	.91		1		
LP-ThR	.35-.64	.52-.69	.79	.32	.47	.41	.60	.64
CT-TMS	.61-.79		.73	.73		1		
CT-ThR	.30-.35	.73-.81	.82	.13	.70	.15	.85	.39
ST-TMS	.86-.88		.90	.90		1		
ST-ThR	.36-.41	.64-.77	.87	.18	.68	.21	79	.46
OE-TMS	.81-.86		.93	.93		1		
OE-ThR	.20-.37	.50-.86	.83	.11	.72	.13	.87	.36
MET-TMS	.53-.85		.89	.89		1		
MET-ThR	.17-.34	.54-.77	.85	.10	.76	.11	.89	.34

NOTE: PR = Problem Recognition; DS = Distress; LP = perceived Legal Pressure; CT = perceived Costs of the Treatment; ST = perceived Suitability of the Treatment; OE = Outcome Expectancy; MET = Motivation to Engage in the Treatment; TMS = TMS-F; ThR = Therapist rating; λ_{Trait} = Standardized estimates of factor loadings on trait-factors; λ_{Method} = Standardized estimates of factor loadings on method-factors.

Treatment, Perceived Suitability of the Treatment, Outcome Expectancy, and Motivation to Engage in the Treatment, which were negative for the TMS-F but positive for the therapist ratings.

The heteromethod correlations are presented in Table 4. All monotrait-heteromethod correlations (at the diagonal), that is, the convergent validities of the TMS-F scales at the level of observed scores, have values of .30 or higher. The discriminant validity of the TMS-F was supported by the fact that in 41 of 42 comparisons, the TMS-F scores had higher correlations with the therapist ratings of the same concept than with those of another concept, that is, the monotrait-heteromethod correlation exceeded the corresponding heterotrait-heteromethod correlations. Only the scores of the Motivation to Engage in the Treatment scale were slightly more correlated with the therapist ratings for Outcome Expectancy than with those for Motivation to Engage in the Treatment.

MTMM Analysis. The CT-C(M-1) model depicted in Figure 2 provided an acceptable fit to the data, χ^2_{MLMV} ($df = 255$, $N = 620$) = 617.3, $p < .001$, CFI = .927, TLI = .920, RMSEA = .048, SRMR = .053. Table 5 provides an overview of the factor loadings of the TMS-F item-parcels and therapist-rating items, and of the reliability, consistency, and method-specificity coefficients of the scales. The latter two coefficients are presented for observed scores and for true scores.

With values between .73 and .93, the reliabilities of all trait-method combinations were satisfactory for the present purpose. As can be seen from the consistency coefficients for the true-score variables, the TMS-F scales explained between 11% (Motivation to Engage in the Treatment) and 41% (Perceived Legal Pressure) of the reliable variance of the therapist ratings. The square roots of the consistency coefficients, which constitute the convergent validities of the TMS-F scales at the level of true scores, ranged between .34 and .64. The high loading of one therapist-rating item on the Perceived Legal Pressure trait factor was probably due to the fact that this item shared an additional *does not apply* response category for voluntary patients with two Perceived Legal Pressure items of the TMS-F. Because of the large number of voluntary patients in the sample (42%), this shared feature may have caused considerable covariance between the two Perceived Legal Pressure measures.

Discussion

The construct validity of the TMS-F was put to a rigorous test applying an MTMM approach with therapist ratings as the criterion method. Overall, the results provide support for the construct validity of the TMS-F scales. With values between .34 and .64, the convergent validity coefficients of the scales were comparable to correlations between self-report measures and ratings by others in other fields of research (Watson, Hubbard, & Wiese, 2000). This result is particularly satisfactory if it is taken into account that many patients were in

the first 3 months of their treatment (see Table 2), and that self-other agreement increases with level of acquaintanceship (Funder & Colvin, 1988), which certainly is lower in many of the present therapist–patient dyads than in the married couples and friendship dyads commonly used in research into self-other agreement (Watson et al., 2000).

In general, the convergence between self-report and observational measures increases when raters are provided with more cues regarding the attribute to be rated (Funder & Colvin, 1997). The relatively strong therapist patient agreement for Problem Recognition, Distress, and Perceived Suitability of the Treatment compared to Perceived Costs of the Treatment, Outcome Expectancy, and Motivation to Engage in the Treatment might reflect that problem recognition and distress are often discussed during treatment sessions, and that patients typically provide ample cues concerning their attitude toward the treatment. In contrast, the patient's perception of the costs of the treatment, outcome expectancy, and commitment to make efforts for the treatment are usually not discussed unless there is a particular reason to do so. The strong agreement concerning the patient's perception of legal pressure may reflect the fact that therapist and patient are both aware of the patient's actual legal status, which underlies the perception of legal pressure.

The discriminant validities of the scales were satisfactory as well. Five of the seven scales had substantially higher correlations with therapist ratings of the same trait (monotrait-heteromethod correlations) than with those of all other traits (heterotrait-heteromethod correlations). The scores of the Motivation to Engage in the Treatment and Perceived Suitability of the Treatment scales had similar correlations with the corresponding therapist ratings as with the therapist ratings of Perceived Suitability of the Treatment and Outcome Expectancy (for Motivation to Engage in the Treatment) and Motivation to Engage in the Treatment (for Perceived Suitability of the Treatment; see Table 4). However, this result must be seen in the light of the high correlations among Outcome Expectancy, Perceived Suitability of the Treatment, and Motivation to Engage in the Treatment that were found for both methods (see Table 3).

The question may rise whether the above conclusions concerning the construct validity of the TMS-F are based on sufficiently firm ground given the unknown validity of the criterion measures. A closer look at the MTMM approach reveals that *prior* validation of the criterion measures is less pivotal for this approach than it is for less sophisticated approaches to test validation, in which conclusions are drawn from *single* correlations and not from *patterns* of correlations. After all, single monotrait-heteromethod correlations (i.e., convergent validity coefficients) can result when both measures reflect the same trait, even if it is not the intended trait. Therefore, it is pivotal to know exactly what the criterion measure measures. In contrast, whole patterns of heteromethod and monomethod correlations, which are obtained from MTMM designs, can only be in accordance with the hypothesized model if all measures involved represent the intended concepts. The fact that in the present study the monotrait-heteromethod correlations exceed almost all corresponding heterotrait-heteromethod correlations (see Table 4) provides evidence for the validity of not only the TMS-F scales but also the therapist-rating scales. As Judd, Smith, and Kidder (1991) put it, "if the theoretically derived predictions turn out to be fulfilled in the data, that constitutes support for all of the components that went into the predictions: the theory as well as the validity of all the measures involved" (p. 57). The fact that the patterns of heterotrait-monomethod correlations for both methods (see Table 3) show many similarities but also differences may be due to the different perspective of therapists and patients. For example, the fact that correlations of Distress with Perceived Costs of the Treatment, Perceived Suitability of the Treatment, Outcome Expectancy, and Motivation to Engage in the Treatment were positive for the therapist ratings but negative for the TMS-F, might reflect that, from the viewpoint of therapists, distress can have a positive connotation as a source of treatment motivation, whereas for patients it is merely a source of suffering.

STUDY 2

Introduction

Because the concept of motivation owes its relevance mainly to the relationship with future behavior, a measure of motivation must be evaluated by its utility to predict behavior. A proximal behavioral criterion for treatment motivation is the patient's effort for the treatment. Therefore, the criterion validity of the TMS-F is evaluated by its power to predict treatment engagement.

In the originally hypothesized model for the TMS-F (see Drieschner et al., 2004), treatment engagement is predicted by the motivation to engage in the treatment, which in turn is predicted by six internal determinants (see Figure 1). However, as reported in Drieschner and Boomsma (2008), Motivation to Engage in the Treatment was virtually unrelated to Perceived Legal Pressure and Distress, and, conditional on the other internal determinant scales, Perceived Suitability of the Treatment was found to be redundant in the prediction of Motivation to Engage in the Treatment despite a substantial correlation (.58). Therefore, and to reduce the complexity of the model, the regression paths of Motivation to Engage in the Treatment on Distress,

TABLE 6
Observed Correlations Among the Sum Scores of the TMS-F Scales and the TER

	PR	*DS*	*LP*	*CT*	*ST*	*OE*	*MET*
PR							
DS	.60						
LP	–.12	–.15					
CT	.08	–.19	–.13				
ST	.32	–.04	.06	.44			
OE	–.04	–.45	.12	.43	.65		
MET	.18	–.16	.06	.36	.50	.61	
TER	.24	–.06	–.08	.28	.47	.39	.47

NOTE: PR = Problem Recognition; DS = Distress; LP = Perceived Legal Pressure; CT = Perceived Costs of the Treatment; ST = Perceived Suitability of the Treatment; OE = Outcome Expectancy; MET = Motivation to Engage in the Treatment; TER = Treatment Engagement Rating scale.

Perceived Legal Pressure, and Perceived Suitability of the Treatment are omitted in the initial model of this study (Model 1). The hypothesis that the internal determinants only have indirect effects on treatment engagement (mediated by Motivation to Engage in the Treatment) is very restrictive because it implies that treatment engagement entirely depends on the explicit behavioral intentions measured by the Motivation to Engage in the Treatment scale. To the extent that the internal determinants influence treatment engagement at a less conscious level, direct effects on treatment engagement could occur. This possibility will be explored unless the fit of Model 1 is fully satisfactory.

Method

Participants. The present sample was a subsample of sample 2 ($N = 376$ from six treatment centers) reported in Drieschner and Boomsma (2008). It consisted of 328 patients from five treatment centers, where treatment engagement ratings could be obtained. Of the 328 patients, 317 were also included in the sample of Study 1. Inclusion criteria were the same as in Study 1.

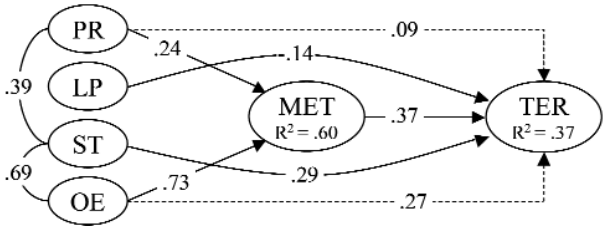
Measures. The TMS-F has been described above. The Treatment Engagement Rating scale (TER) is a Dutch 21-item therapist-rating instrument for the treatment engagement of patients in forensic outpatient treatment (Drieschner & Boomsma, in press). It was developed on the basis of the literature about necessary or desirable behavior of patients in psychological treatment and interviews with experienced therapists from the field of forensic outpatient treatment. The TER addresses nine components of treatment engagement, called session attendance and punctuality (2 items), making sacrifices (3), openness (2), efforts to change problem behavior (2), goal-focus (2), efforts for improvement of one's social-economic situation (5), constructive use of treatment sessions (3), reflecting on treatment content between sessions (1), and overall treatment engagement (1). With the exception of session attendance and punctuality, all items have anchored 5-point response scales. The meanings of the midpoint and the two endpoints of the response scales are defined by brief descriptions of behavior for which the score is appropriate. The scores of the items for session attendance and punctuality represent the proportions of sessions in the rating period for which the patient did not turn up and was late, respectively. These proportions are recoded into 5-point scales. The means of the item scores within each component constitute the component scores, and the sum of the nine component scores constitutes the TER total score. All raters received 90 minutes of rater training, in which each item was discussed on the basis of practical examples, and information about common rater biases, such as halo and leniency bias, was provided. A psychometric evaluation of the TER is reported in Drieschner and Boomsma (in press). In summary, the nine component scores were found to constitute a homogeneous scale, as was indicated by single-factor confirmative factor analysis. Cronbach's alpha of the total score computed from the nine component scores was .93. The interrater reliability of the TER total score in a subsample of the present sample ($N = 99$) was .76. The correlation of the TER total scores with the component scores of session attendance and punctuality was .53.

Procedure. The TMS-F was administered as described in Study 1. Between 8 and 13 weeks later, therapists, who were kept blind for the TMS-F scores, rated the patients' efforts during the previous eight weeks with the TER. In most cases (60.7%), only one TER rating per patient was provided. When several therapists were involved in the treatment, two (32.3%) or more (7%) TER scores were available, and the mean of these scores was used in the analyses.

Data Analysis

First, Pearson correlations between TMS-F scores and TER total scores were computed. Of primary interest was the correlation of the Motivation to Engage in the Treatment with the TER total score, which constitutes the predictive validity of the Motivation to Engage in the Treatment scale at the level of observed scores. Next, structural equation modeling was applied to test the initial model (Model 1), the structural part of which is presented in Figure 1 (unbroken lines). Each of the six internal determinant factors and the TER factor were linked to 9 and Motivation to Engage in the Treatment to 16 observed variables. Finally, depending on the fit of

FIGURE 3
Path Diagram of the Structural Part of Model 3 with Standardized Parameter Estimates



Note: All values represent standardized parameter estimates. Broken lines represent indirect effects on TER mediated by MET. PR = Problem Recognition; DS = Distress; LP = Perceived Legal Pressure; CT = Perceived Costs of the Treatment; ST = Perceived Suitability of the Treatment; OE = Outcome Expectancy; MET = Motivation to Engage in the Treatment; TER = Treatment Engagement Rating scale.

Model 1, modified models with additional direct effects of the internal determinants on TER are fitted to the data.

As in Study 1, MLMV estimation as implemented in Mplus was applied. Model fit was evaluated by Hu and Bentler's (1999) combinatory cutoff criteria. Because items instead of item-parcels were used as observed variables, which resulted in lower factor loadings, we followed the recommendation of Beauducel and Wittmann (2005) to give priority to the combination of the SRMR and RMSEA. This recommendation was based on the finding that in models with low or moderate factor loadings, incremental fit indices such as the TLI and the CFI penalize unspecified small secondary factor loadings (i.e., loadings on unintended factors) to a degree that such models "would only have a chance to be accepted when incremental fit indexes and the GFI are not used for model evaluation" (p. 70). In multidimensional questionnaires, secondary factor loadings are hardly avoidable, if only because of formal or linguistic features shared by items of different scales (see Angleitner, John, & Löhr, 1986). Analyses reported in Drieschner and Boomsma (2008) supported that the low TLI and CFI values reflect comparatively small secondary factor loadings of little practical importance.

Results

Correlations Among Observed Scores. The Pearson correlations among the observed scores of the TMS-F scales and the TER are summarized in Table 6. The correlation between Motivation to Engage in the Treatment and TER was .47, corresponding with 22.1% of the variance of TER explained by the Motivation to Engage in the Treatment. TER was similarly correlated with Perceived Suitability of the Treatment and Outcome Expectancy. The correlations of TER with Perceived Costs of the Treatment and Problem Recognition were moderate, those with Distress and Perceived Legal Pressure nonsignificant.

Structural Equation Modeling. The fit of Model 1 (see Figure 1, unbroken lines), χ^2_{MLMV} (df = 224, N = 328) = 400.8, $p < .001$, CFI = .810, TLI = .808, RMSEA = .049, SRMR = .086, was acceptable according to the combination of RMSEA < .06 and SRMR < .10. As argued in the Method section, the low CFI and TLI values were likely to reflect small unspecified secondary factor loadings. An additional reason for the low CFI and TLI values may be an overrejection tendency of these indices when the assumption of multivariate normality is violated (West, Finch, & Curran, 1995), especially in large models (Gerbing & Anderson, 1993; Hutchinson & Olnos, 1998; Kenny & McCoach, 2003).

Modification indices suggested two additional direct effects of Perceived Suitability of the Treatment and Perceived Legal Pressure on TER. When these paths were added (Model 2), the fit improved significantly, $\Delta\chi^2$ (df = 2) = 23.3, $p < .001$. Distress had no significant direct or indirect effect on either Motivation to Engage in the Treatment or TER. The same applied for Perceived Costs of the Treatment in spite of substantial correlations of .38 and .31 with the Motivation to Engage in the Treatment and TER, respectively. Therefore, both scales were removed. The resulting more parsimonious model (Model 3) was taken as the final model.

Model 3 provided an acceptable fit to the data, χ^2_{MLMV} (196, N = 328) = 353.1, $p < .001$, CFI = .845, TLI = .842, RMSEA = .049, SRMR = .079. The standardized regression weights in the structural part of Model 3 are presented in Figure 3. Together, Motivation to Engage in the Treatment, Perceived Suitability of the Treatment and Perceived Legal Pressure accounted for 37.1% of the variance of TER. Problem Recognition and Outcome Expectancy had significant indirect effects on TER, mediated by Motivation to Engage in the Treatment. The latent correlation between Motivation to Engage in the Treatment and TER, which constitutes the predictive validity coefficient of the Motivation to Engage in the Treatment scale at the level of true scores, was .55. This corresponds with 29.9% of the variance of the TER explained by Motivation to Engage in the Treatment alone. The latent correlations of TER with Perceived Suitability of the Treatment and Outcome Expectancy were only slightly lower, with .51 and .47, respectively. Those with Problem Recognition and Perceived Legal Pressure were .20, and –.14, respectively.

Discussion

The criterion validity of the TMS-F was evaluated by its utility to predict Treatment Engagement. The main conclusion of the study is that TMS-F scores, in particular those of the Motivation to Engage in the Treatment scale, predict treatment engagement to a substantial degree. A latent correlation of .55 between a self-report measure and a rating of subsequent behavior, and 37.1% of the variance of future behavior explained by self-report scores compare favorably with values found in other fields of research (Hagger, Chatzisarantis, & Biddle, 2002; Kraus, 1995). This result is especially encouraging because self-report of offenders is often regarded as untrustworthy (e.g., Lanyon, 2001; Nieberding, Moore, & Dematatis, 2002).

The study yielded additional remarkable results. The fact that distress, as measured by the Distress scale, was virtually unrelated to treatment engagement is noteworthy because lack of distress is often regarded as a major reason for deficient treatment motivation of offenders, and creating a sense of distress is considered important to replace external by internal treatment motivation (e.g., Hemphill & Hart, 2002). It is possible that distress is motivating for some but demoralizing for other patients. In that case, both effects might balance each other at the level of groups, although distress is relevant for the motivation of each individual patient.

Second, a *lower* level of legal pressure, as measured by the Perceived Legal Pressure scale, predicted *more* treatment engagement, although the association was weak. In a secondary analysis with the data of the court mandated patients only, exactly the same association ($r = -.08$) was found. This is remarkable because in the field of offender treatment "perceptions of threat [. . .] are among the commonly identified mediators between actual sanctions and behavior" (Maxwell, 2000, p. 544). It is possible that the perception of legal pressure motivates offenders to *enter* a treatment but not to make efforts for the treatment, as long as they believe that this will have no legal consequences. Interestingly, only 46% of the *court mandated* patients in Studies 1 and 2 agreed with the item, *It might have legal consequences if therapists think that I make too little effort for the treatment*. This is in accordance with our observation that patients were rarely excluded from the treatment because of nonattendance or lack of treatment engagement, unless they violated basic rules such as the ban on violence or drug use. It should be noted that the present results do not justify the final conclusion that legal pressure has a negative effect on treatment engagement. It is possible that judges impose more legal pressure if they think an offender has less intrinsic treatment motivation. In that case, the legal pressure might enhance the treatment engagement of individual patients despite the negative correlation at the level of the group.

Third, it is interesting that Perceived Suitability of the Treatment has a *direct* effect on TER. Recall that one of the aspects addressed by the Perceived Suitability of the Treatment scale is the satisfaction with the therapist and the therapeutic relationship (see Table 1). It is possible that the Perceived Suitability of the Treatment scores partly reflect an emotional response to the therapist, which relates to the behavior of treatment engagement at a less conscious level than the behavioral intentions assessed by the Motivation to Engage in the Treatment scale.

Finally, it is notable that Motivation to Engage in the Treatment scores predicted treatment engagement across different stages of the treatment. This is important because the treatment motivation of patients is a dynamic variable, which deserves the continuous attention of therapists (Miller, 1985). The present results suggest that the TMS-F can be used in longitudinal research into the waxing and waning of the treatment motivation or for evaluating the effects of motivational interventions.

Because the final model partly resulted from data-driven modifications of the initial model, some capitalization on chance may have occurred. However, this does not apply to the most important result of this study, the predicted substantial correlation between the patient's motivation to engage in the treatment, as measured by the Motivation to Engage in the Treatment scale, and the subsequent behavior of treatment engagement.

GENERAL DISCUSSION

Rosenbaum and Horowitz (1983) noted with respect to the concept of treatment motivation that "to prove useful in psychotherapy research, it must define some restricted, clearly defined conceptual domain which provides some predictive power" (p. 351). The general conclusion from the present studies is that the TMS-F scales and their underlying concepts satisfy both requirements. The TMS-F scales were found to measure what they were supposed to measure (Study 1), and to predict the behavior they were supposed to predict on theoretical grounds (Study 2). In common terminology, Study 1 supports the construct validity, Study 2 the predictive validity of the TMS-F.

Given these results and the earlier evidence for the factorial validity of the TMS-F (Drieschner & Boomsma, 2008), TMS-F scores can now be interpreted with some confidence. Doing so, it seems that treatment efforts mostly depend on the patients' appraisal of the treatment (i.e., Perceived Suitability of the Treatment and Outcome Expectancy), and only marginally on factors that exert

pressure to enter the treatment (i.e., Distress, Perceived Legal Pressure, and Problem Recognition). With respect to legal pressure, our results are in accordance with an earlier finding that legal pressure predicts retention but not treatment engagement (Knight, Hiller, Broome, & Simpson, 2000). The importance of the patient's appraisal of the treatment suggests that delivering forensic outpatient treatment in a convincing and encouraging way is important for the patients' treatment responsivity.

Although the scales for Distress and Perceived Costs of the Treatment did not make an incremental contribution to the prediction of either Motivation to Engage in the Treatment or treatment engagement, there are reasons to retain these scales in the TMS-F. Most importantly, both scales are valid measures for their respective concepts, as was shown in Study 1. Given the present evidence, it would be premature to abolish valid measures of variables that are widely regarded as relevant. Because the Perceived Costs of the Treatment scale predicts both Motivation to Engage in the Treatment and TER if taken alone, it could be useful in individual assessment situations. For example, when a patient repeatedly misses treatment sessions in spite of satisfactory Perceived Suitability of the Treatment and Outcome Expectancy scores, it would be informative to find a Perceived Costs of the Treatment score indicating that the treatment is perceived as highly aversive. The Distress scale may have predictive utility in interaction with other variables, such as the source of distress, treatment history, or kind of psychopathology. For example, distress may affect the patient's treatment motivation differently if it arises from an addiction with a history of unsuccessful treatments, or from a comparatively new problem and worries about its escalation in the future.

A validity issue that was addressed only indirectly is the influence of social desirability response bias, which is generally regarded as a major threat to the validity of self-report in forensic psychiatry. The substantial correlations between TMS-F scores and independent therapist ratings demonstrate that the validity of the TMS-F is not seriously eroded by this response bias. A more direct investigation of the influence of social desirability on the TMS-F scores will be reported in another publication (see also Drieschner, 2005). Briefly summarized, the factorial structure of the TMS-F was found to be invariant across respondents with above median and below median scores on a validated social desirability scale. Moreover, partialing out the effect of social desirability did not enhance the correlations between TMS-F scores and the scores on observational criterion measures, which themselves were uncorrelated with the scores of the social desirability scale. Higher correlations would have been expected if a substantial proportion of the variance of the TMS-F scores would reflect social desirability response bias.

The validation of a psychological measurement instrument is an ongoing process. Several validity issues should be addressed by further research. First, it should be investigated to what extent the factorial validity, construct validity, and predictive validity of the TMS-F generalize to various subgroups of the target population, such as various types of offenders, substance abusers, psychotic individuals, and individuals with a non-Western cultural background. Second, the construct validity of the scales should be further investigated with other criterion measures than those employed in Study 1. Third, the predictive utility of the TMS-F should be investigated with more distant criteria, such as dropout or reoffending. Finally, because there is always some ambiguity in the interpretation of correlation coefficients, validation research that does not rely on correlations would be desirable. For example, if it could be shown that Perceived Legal Pressure scores increase after an intervention to enhance the perception of sanction threat, this would be evidence for the validity of the Perceived Legal Pressure scale. In the same vein, interventions designed to enhance optimism concerning treatment outcomes, or to create commitment for treatment engagement could be used to validate the Outcome Expectancy and Motivation to Engage in the Treatment scales, respectively. Although additional validation research is desirable, the present evidence for the construct validity and predictive validity of the TMS-F, together with the available evidence for the factorial validity of the instrument, the homogeneity of the scales, and the stability of these properties across independent samples (Drieschner & Boomsma, 2008), are encouraging and suggest that the TMS-F can be a valuable instrument for clinical purposes and research in the field of forensic outpatient treatment.

REFERENCES

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52*, 7-27.

Angleitner, A., John, O. P., & Löhr, F. J. (1986). It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement* (pp. 61-108). Berlin, Germany: Springer.

Barbaree, H. E. (2005). Psychopathy, treatment behavior, and recidivism. *Journal of Interpersonal Violence, 20*, 1115-1131.

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*, 41-75.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollins.

Drieschner, K. H. (2005). *Measuring treatment motivation and treatment engagement in forensic psychiatric outpatient treatment: Development of two instruments*. Unpublished Dissertation, Radboud University, Nijmegen, Netherlands.

Drieschner, K. H., & Boomsma, A. (2008). *The Treatment Motivation Scales for forensic outpatient treatment (TMS-F): Construction and psychometric evaluation*. *Assessment*, [XX, xx-xx].

Drieschner, K. H., & Boomsma, A. (in press). The Treatment Engagement Rating Scale (TER) for outpatient offender treatment: description, psychometric properties, and norms. *Psychology, Crime & Law*.

Drieschner, K. H., Lammers, S. M. M., & Van der Staak, C. P. F. (2004). Treatment motivation: An attempt for clarification of an ambiguous concept. *Clinical Psychology Review, 23*, 1115-1137.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*, 241-261.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods, 8*, 38-60.

Falshaw, L., Friendship, C., Travers, R., & Nugent, F. (2003). *Searching for what works: An evaluation of cognitive skills programmes*. London: Home Office.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149-158.

Funder, D. C., & Colvin, C. R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 656-672). San Diego, CA: Academic Press.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.

Hagger, M. S., Chatzisarantis, N. L. D., & Biddle, S. J. H. (2002). A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: Predictive validity and the contribution of additional variables. *Journal of Sport and Exercise Psychology, 24*, 3-31.

Hemphill, J. F., & Hart, S. D. (2002). Motivating the unmotivated: Psychopathy, treatment, and change. In M. McMurran (Ed.), *Motivating offenders to change: A guide to enhancing engagement in therapy* (pp. 193-221). New York: John Wiley.

Howells, K., & Day, A. (2007). Readiness for treatment in high risk offenders with personality disorders. *Psychology, Crime & Law, 13*, 47-56.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hutchinson, S. R., & Olnos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling, 5*, 344-364.

Jackson, D. N. (1971). The dynamics of structured personality tests. *Psychological Review*, *78*, 229-248.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). New York: Holt, Rinehart & Winston.

Kenny, D. A., & Kashy, D. A. (1992). The analysis of the multitrait-multimethod matrix by confirmative factor analysis. *Psychological Bulletin, 112*, 165-172.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*, 333-351.

Knight, K., Hiller, M. L., Broome, K. M., & Simpson, D. D. (2000). Legal pressure, treatment readiness, and engagement in long-term residential programs. *Journal of Offender Rehabilitation, 31*, 101-115.

Knight, K., Holcom, M., & Simpson, D. D. (1994). *TCU Psychosocial Functioning and Motivation Scales: Manual on psychometric properties*. Fort Worth: Texas Christian University—Institute of Behavioral Research.

Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin, 21*, 58-75.

Lanyon, R. I. (2001). Psychological assessment procedures in sex offending. *Professional Psychology: Research and Practice, 32*, 253-260.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighting the merits. *Structural Equation Modeling, 9*, 151-173.

Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & Van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's sex offender treatment and evaluation project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment, 17*, 79-107.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177-198). Thousand Oaks, CA: Sage.

Maxwell, S. R. (2000). Sanction threats in court-ordered programs: Examining their effects on offenders mandated into drug treatment. *Crime & Delinquency, 46*, 542-563.

McConnaughy, E. A., Prochaska, J. O., & Velicer, W. F. (1983). Stages of change in psychotherapy: Measurement and sample profiles. *Psychotherapy: Theory, Research and Practice, 20*, 368-375.

McGuire, J. (Ed.). (1995). *What works: Reducing reoffending: Guidelines from research and practice*. Chichester, UK: Wiley.

McMurran, M. (Ed.). (2002). *Motivating offenders to change: A Guide to enhancing engagement in therapy*. New York: John Wiley.

Miller, W. R. (1985). Motivation for treatment: A review with special emphasis on alcoholism. *Psychological Bulletin, 98*, 84-107.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide (version 3)*. Los Angeles: Muthén & Muthén.

Nieberding, R. J., Moore, J. T., & Dematatis, A. P. (2002). Psychological assessment of forensic psychiatric outpatients. *International Journal of Offender Therapy and Comparative Criminology, 46*, 350-363.

Plemper, E. (2001). *Ambulante forensische psychiatrie: Onderzoek forensisch psychiatrische poli-en dagklinieken* [Forensic outpatient psychiatry: Investigation of forensic psychiatric outpatient treatment centers]. Utrecht, Netherlands: Verwey-Jonker Instituut.

Rosenbaum, R. L., & Horowitz, M. J. (1983). Motivation for psychotherapy: A factorial and conceptual analysis. *Psychotherapy: Theory, Research and Practice, 20*, 346-354.

Salize, H. J., & Dressing, H. (Eds.). (2005). Placement and treatment of mentally disordered offenders—legislation and practice in the European Union. Lengerich, FRG: Pabst Science.

Sellen, J. L., McMurran, M. M., Cox, W. M., Theodosi, E., & Klinger, E. (2006). The Personal Concerns Inventory (Offender Adaptation): Measuring and enhancing motivation to change. *International Journal of Offender Therapy and Comparative Criminology, 50*, 294-305.

Serin, R., & Kennedy, S. (1997). *Treatment readiness and responsivity: Contributing to effective correctional programming*. Ottawa: Research Branch, Correctional Services of Canada.

Sijtsma, K., & Van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research, 38*, 505-528.

Tierney, D. W., & McCabe, M. P. (2002). Motivation for behavior change among sex offenders: A review of the literature. *Clinical Psychology Review, 22*, 113-129.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78*, 546-558.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (version 2)*. Burnaby, Canada: Simon Fraser University.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Williamson, P., Day, A., Howells, K., Bubner, S., & Jauncey, S. (2003). Assessing offenders' readiness to change problems with anger. *Psychology, Crime & Law, 9*, 295-307.